Fitness Optimizing Theory Can be Misleading when Applied to Human Behavior

Bjørn Grinde*

Norwegian Institute of Public Health, Oslo, Norway

Abstract: The mathematical approach to understand the genetic impact on human behavior, e.g., using game theory, tends to assume that evolution creates optimal fitness. However, evolution does not necessarily create optimal behavioral patterns, particularly not when the environment is rapidly changing. Consequently some of the inferences based on game theory are misleading. More specifically, certain aspects of current human behavior may be better understood in terms of emotions and motivational systems. The present text is an attempt to discuss possible pitfalls of fitness optimalisation theory by looking at patterns of behavior, such as cooperativity, sex and art, where this theory alone may not give satisfactory explanations.

INTRODUCTION

The use of mathematical modeling to understand the evolution of behavioral traits goes back to the sixties [1, 2]. Borrowing game theory from the field of economics, these authors developed the concept of evolutionary stable strategy (ESS). ESS deals with the basic tenets of evolutionary theory in a formalized way. It thus allows for mathematical modeling of human behavioral choices based on the premise that we are primed by evolution to optimize our fitness, i.e., to choose the options that are likely to bring the more genes into coming generations. An evolutionary strategy is stable if it performs accordingly.

The use of game theory has improved our understanding of human behavior appreciably, e.g., by explaining sociability as rooted in evolutionary strategies such as inclusive fitness and reciprocal altruism. However, the ensuing tendency to explain traits in terms of fitness optimizing may occasionally lead astray, primarily because evolution not always creates optimal functions or strategies.

There are several constraints that explain why evolution in general does not always lead to optimal adaptations. Briefly, observed behavior may be suboptimal for the following reasons: One, optimality is generally not required, all that is required is for the organism to be capable of survival and reproduction. Two, optimality is often not achievable because the process of evolution is restricted in various ways, including the demand for a scaffold as a starting point for molding changes. Three, in cases where the environment differs from what the organism is genetically adapted to, behavior does not reflect an adaptation to the environment at hand. The latter is the most relevant point for the present discussion.

The above limitations are usually dealt with in the literature applying game theory and ESS to animal behavior, but applying game theory to human behavior is somewhat more problematic. There are several examples where fitness optimizing theory leads to inadequate explanations for observed behavior.

The foundations for human emotions and motivations are formed by the evolutionary process, and are thus not necessarily designed for optimal reproductive success in present society. Rather than considering present behavior to be fitness optimizing, one should try to understand how our these aspects of the brain were molded by evolution. That is, knowledge as to how archaic feelings and motivational systems influence human actions should help improve explanations. Although these predispositions evolved to improve fitness, their effect on present behavior is not necessarily fitness enhancing.

In the present text, the terms emotions, feelings and motivations will be used as outlined in [3, 4]. Emotions are reasonably distinct functions laid down in the brain in order to influence behavior, but not directed primarily at satisfying bodily needs. The word "feeling" is broader and includes both emotions and functions such as thirst and hunger. The term "motivations" focuses more on the neurological instruments within the brain that help direct behavior, i.e., the brain functions that compel us to choose an action. Motivational systems can thus be considered as neurological elements, associated with the various types of feelings, and having the capacity to influence behavior [5]. In other words, feelings include (or are connected with) motivational systems. The latter may informally be referred to as brain rewards and penalties.

MOTIVATIONAL SYSTEMS: BRAIN REWARDS AND PENALTIES

A main purpose of the brain is to direct behavior. This function is cared for by generating a register of emotions and feelings such as fear, anger, guilt, empathy, pleasure and pain. The corresponding motivational systems are typically designed by evolution for two opposing purposes: One is to instigate or promote relevant behavior, the other to teach the individual to avoid particular situations or discourage certain actions [6]. This duality of objective implies that, as a gross simplification, emotions and feelings can be classified as either positive or negative experiences. In order to reflect this

*Address correspondence to this author at the Norwegian Institute of Public Health, P.O. Box 4404 Nydalen, 0403 Oslo, Norway; E-mail: bjgr@fhi.no

duality, positive experiences may be referred to as brain rewards, and the negative experiences as brain penalties.

It should be noted that when used this way, the terms rewards and penalties cover more or less all emotions and feelings, not just the distinct pleasure and pain circuits of the brain that these terms occasionally refer to. Although the present use represents a simplification, e.g., what is considered rewarding for one person is not necessarily rewarding for others, they do serve the purpose of pointing out that emotions and feelings typically either encourage or discourage behavior.

Motivational systems were designed to improve fitness in the environment of evolutionary adaptation (EEA); but, as pointed out in the Introduction, that does not necessarily lead to optimal behavior. The point is particularly relevant as the conditions of modern societies in many ways are drastically changed compared to the human EEA.

Another relevant point is that emotional instigations, in the form of brain rewards and penalties, may be more pervasive in humans as compared to animals. The human lineage presumably saw an expansion of conscious processes with a concomitant increase in free will. As a consequence, evolution may have relied to a greater extent on motivational systems (rather than instincts) to direct human behavior towards what would be in the interest of the genes. In other words, more powerful feelings may have been required to counteract our, potentially misleading, conscious thinking in situations where our animal ancestors relied to a greater extent on instinctive tendencies. Moreover, key aspects of human behavior, such as cooperativity, most likely evolved late. This assumption is based on the observation that most apes, with the exception of chimpanzees, do not cooperate appreciably within genders, suggesting that the evolution of social behavior in our lineage is primarily associated with the last 5 million years. Presumably behavioral modifications were founded to a greater extent on motivations rather than on instincts during this stage of human evolution.

As suggested in the above paragraph, motivational systems may have a greater impact on humans as compared to animals. Consequently one should be particularly careful when using a fitness optimizing approach to understand human behavior in general and social behavior in particular.

COOPERATIVITY

A characteristic feature of human behavioral biology is that our brains are adapted to social life. Mathematical modeling and game theory have been particularly popular when it comes to understanding the evolution of human altruism and cooperativity (for recent reviews, see [7, 8]). However, although the basic tenets have proven useful, when applied to explain certain observations, such as the choices made in dilemma games, the inferences may be misleading.

Various game theory inspired dilemma type experiments have been designed to probe cooperative versus selfish behavior, and to investigate behavior related to the notion of fairness, as reflected in direct sanctions or third party punishment [9-13]. See also [14] for a recent attempt to explain the connection between punishment and altruism.

An illustrative example of dilemma type experiments is offered in [10]. The authors assigned students pair wise to be either "investors" or "trustees", and gave the investors 10 monetary units (MU). The investors had the option to transfer from 0 to 10 MU to the trustees. The trustees received, from the organizer, three times the amount transferred and could hand back whatever they found fair. So if an investor transferred the entire 10 MU, the corresponding trustee would receive 30, sitting on a total of 40. If the trustee chose to return 20, the investor would end up with twice the sum he started with; if he returned none, the investor would end up with nothing.

Altruism was demonstrated by the observation that most of the trustees chose to return some of the money. However, the more intriguing results came when the experimenter gave the investor the option of imposing a fine of 4 MU in case the trustee would return less than the investor stated that he expected to get. In this setting, those who let the trustee know there would be no fine – whatever he returned – received the most money, suggesting that the trustees rewarded those who did not use threats. Moreover, in a second round of experiments, people levied a fine even if they knew that the expected return would improve by not doing so. As the interactions were totally anonymous, and one time only, neither direct nor indirect reciprocity applied, i.e., one could not gain anything by being nice or by building a reputation.

The observed behavior is not in concordance with game theory in that the participants did not act in a manner that would give them the most benefit (in this case money) in return. However, the observations do make sense in view of motivational systems related to the idea of fairness and to the feeling of empathy. We tend to enjoy being nice to each others, while we dislike the use of unwarranted threats. In a tribal setting, the underlying motivational system presumably had survival value in that they improved group performance. Furthermore, in the tribal setting there was no need to limit the behavior to situations involving kin or comrades because one rarely interacted with strangers. These motivational systems would be expected to influence the options chosen in the above game situation, regardless of whether the participants knew each other or expected to meet again.

As pointed out in [15], human cooperation often defies rational-choice theory. For example, the generosity towards strangers appears puzzling, as does the willingness to levy third party punishment even if the consequences for self are negative. The enigma relates both to the above mentioned laboratory type experiments, and to the observed behavior of everyday life, such as the readiness to donate money to strangers, and the keenness to adopt unrelated children. Game theory encounter problems explaining why people cooperate with genetically unrelated strangers they never met, or with people that they meet once but are unlikely to ever see again, as such actions are unlikely to benefit their fitness. Emotional instigations in the form of brain rewards and penalties seem to improve our explanatory power.

The brain rewards associated with social situations, such as those evoked in the dilemma games, evolved under quite different circumstances. The human EEA presumably implied small groups of individuals with strong and lasting bonds, and often genetic relatedness. Most interactions involved individuals with long-term commitments, either within a tribe or between cooperating tribes. Obviously hostile encounters occurred, but then the opponents would be clearly defined as enemies; and the rules of behavior associated with cooperativity would not apply. The important point is that evolution did

not adapt behavior to neutral, one-time encounters with strangers, because such situations were uncommon. Instead we are primed to deal with any person as either a friend or a foe, and motivated by brain rewards and penalties to act accordingly.

During the formative period of human evolution, it was not possible to donate resources to far away people. Moreover, levied punishment would be for the benefit of the group and thus potentially also to the benefit of the individual carrying it out, at least to the extent that it might improve reputation. Anonymity was more often not feasible. In other words, there was no obvious evolutionary cost of letting the general interactions with none-enemies be based on empathy and cooperativity. The choices we observe today are influenced by motivational systems installed under these circumstances. Consequently, it feels good to give, even to total strangers, and it feels good to contribute to the enforcement of acceptable group behavior in the form of punishing deviants.

The underlying principles of game theory – such as kin selection, reciprocal altruism, indirect reciprocity/reputation, and, possibly, group selection – are needed to explain why these behaviors, and concomitant motivational systems, were selected for in the first place, but one should be careful when applying these principles to explain current behavior.

SEXUAL BEHAVIOR

The relevance of considering brain rewards, and not just fitness optimizing, is particularly striking when examining sexual behavior.

A key observation in this respect is that humans eagerly participate in sexual activity in the presence of contraceptives. As this does not appear to be an evolutionary stable strategy, one is unlikely to find an explanation based solely on fitness optimisation. However, the fact that sexual activity is associated with ample brain rewards accounts for the observations. In other words, people take part in non-reproductive sex primarily because they consider it a rewarding experience. Contraceptives were unavailable during the EEA, thus there was no selection aimed at avoiding their use.

It might be argued that non-reproductive sexual behavior is fitness enhancing because it promotes bonding. I.e., the cost of non-reproductive sex is countered by an increasing cohesiveness of the couple and subsequent improved infant care. Although the point most likely is relevant, it does not fully explain observed behavior. For one, pair-bonding intentions are not required for participation, particularly not on behalf of the male; and two, self-stimulatory sex is common.

A parsimonious model would be to suggest that human reproductive biology is based on a sexual drive rather than on an urge to create offspring. Attachment to the baby, and concomitant caring behavior, develop primarily after birth. In the EEA this strategy was evolutionary stable, that is, it worked according to intentions. As non-reproductive sex became more of an option, the observed behavior is no longer evolutionary optimal. Thus current behavior is best explained by implicating motivational systems, in the form of brain rewards, molded during the EEA.

The discussion on sexual behavior may seem obvious. The point is that in the more complex situation of social behavior, as reflected in the above discussion on cooperativity, the impact of motivational systems is less clear. The example of sexual behavior is meant to emphasize a particular way of thinking that may be useful when analyzing a variety of human activities.

ART AND MUSIC

There have been several attempts at explaining the human fascination for aesthetics in terms of fitness optimisation [16-18]. Most accounts focus on sexual selection, that is, both visual art and music evolved as a display to attract partners. It is definitely conceivable that singing or self decoration could improve attractiveness; as would the production of aesthetic objects or the handling of a musical instrument. Thus sexual display may very well have played a part in the evolution of artistic propensities. Yet, implicating brain rewards may improve our explanation as to the present preoccupation with art.

It is noteworthy that both visual and auditory aesthetics are eagerly participated in outside the setting of reproductive behavior, e.g., alone, or by infants and the elderly. Even some animals, particularly chimpanzees and gorillas, appear to delight in painting in context that is unlikely to imply any sexual display; moreover, the animals include aesthetic elements recognized by human observers [19].

Another problematic issue is that the authors who point to sexual selection to explain visual aesthetic behavior also tend to state that this sort of behavior originated approximately 40 000 years ago. It is generally agreed that our species, i.e., modern humans, is at least 200 000 years old, that the last common ancestor of present humans most likely dates back more than 100 000 years, and that the first branching leading to the present subpopulations occurred some 100 000 years ago (reviewed in [20, 21]). In order to explain that aesthetic behavior evolved 40 000 thousand years ago, and at the same time explain that it permeates all existing cultures, one would need to imply extensive convergent evolution. This does not seem likely.

As previously pointed out [22], the concept of brain rewards can help account for several of the features associated with visual aesthetics. Briefly, aesthetics may be about recognizing visual cues that, for various reasons, evolved to be associated with positive feelings.

One example is a delight in color, and particularly what is considered to be "warm" colors such as red, orange and yellow. Higher primates are among the few mammals with trichromatic vision, presumably a consequence of the importance of color vision for finding food [23]. It is paramount for fruit eaters to recognize ripe fruits, which typically are colored in shades from yellow to red. Observing such colors (or any color that stands out in contrast) was presumably connected with positive feelings in order to encourage our ancestors to find and gather valuable nourishment.

Several other aspects of visual stimuli, such as impressions associated with curiosity, with balance and symmetry, or with depth and movement, might also induce positive feelings.

The central idea is that art may have developed as a way of catering to relevant brain rewards induced by visual cues. Moreover, art may involve a propensity to transform these cues into superstimuli. The notion is supported by the dictionary definition of beauty as "a quality that gives pleasure to the mind".

As to music, in addition to sexual display theory, a likely association with language has been pointed out (for a recent review, see [24]). The main point being that our preoccupation with music must have something to do with the obvious survival value of advanced oral communication.

The connection between language and our passion for music may be explained in terms of inherent reward mechanisms. Play behavior is fun because the brain offers rewards for activities that improve our physical and mental skills. Similarly, song and music might be rewarding because they involve practicing of our oratory and/or auditory skills. The qualities of music appreciated (such as purity of tone, rhythm, and complexity) tend to be those associated with optimal communication. As previously discussed [25], music most likely has the capacity to cater to several different types of reward mechanisms. Furthermore, as in the case of visual art, music may have developed ways of stimulating our innate, sound-related reward mechanisms in an exaggerated way. Thus present music may elicit feelings that are stronger than anything triggered by sound in the Paleolithic period.

It should be pointed out that the use of brain rewards in explaining visual or auditory aesthetic behavior does not exclude that other factors, such as status and sexual selection, has played a role. The important point is that present behavior is best explained by including various motivational instigations that most likely date back prior to the emergence of modern humans. The strong impact of music and art in modern societies may be partly a cultural phenomenon, and partly a question of the invention of tools that allow us to fine-tune and exaggerate reward related stimuli.

CONCLUSIONS

The main purpose of the present text is to point out that motivational systems should be taken into account when explaining current human behavior. A too strong focus on mathematical modeling based on fitness optimizing will in many cases not lead to satisfactory explanations for the observed behavior. The present review includes only a few examples of behavioral propensities where this notion ought to be considered.

The idea of evolution as a fitness enhancing process holds, but there are important limitations. For example, it is particularly essential to realize that when the conditions change, the behavioral propensities created by evolution may be far from optimal.

As to altruistic behavior, it has been shown that in dilemma games mutual cooperation is associated with activation of brain areas linked to reward processing [26]; and that motivational systems underlie our willingness to punish others [27]. These reports substantiate the idea that brain rewards help explain the choices observed both in real life and in experimental settings. Human social behavior evolved by installing motivational systems associated with notions such as fairness, reputation, good deeds and trust. These feelings will influence behavior even if the situation is considerably different from the evolutionary context. It feels good (which means that motivational systems in the form of brain rewards are activated) to give, to trust, and to be trusted.

REFERENCES

[1]   Hamilton WD. The genetical evolution of social behaviour. J Theor Biol 1964; 7: 1-52.
[2]   Maynard Smith J. Mathematical Ideas in Biology. Cambridge University Press: Cambridge; 1968.
[3]   Panksepp J. Affective Neuroscience. Oxford University Press: Oxford; 1998.
[4]   Panksepp J. Affective consciousness: Core emotional feelings in animals and humans. Conscious Cogn 2005; 14: 30-80.
[5]   Bouton ME. Learning and Behavior: A Contemporary Synthesis. Sinauer Associates: Sunderland; 2006.
[6]   Keller MC, Nesse RM. Is low mood an adaptation? Evidence for subtypes with symptoms that match precipitants. J Affect Dis 2005; 86: 27-35.
[7]   Fehr E, Fischbacher U. The nature of human altruism. Nature 2003; 425: 785-91.
[8]   Lehmann L, Keller L. The evolution of cooperation and altruism – a general framework and a classification of models. J Evol Biol 2006; 19: 1365-76.
[9]   Fehr E, Gächter S. Altruistic punishment in humans. Nature 2002; 415: 137-40.
[10]   Fehr E, Rockenbach B. Detrimental effects of sanctions on human altruism. Nature 2003; 422: 137-40.
[11]   Fehr E, Fischbacher U. Third-party punishment and social norms. Evol Hum Behav 2004; 25: 63-87.
[12]   Gürerk Ö, Irlensbusch B, Rockenbach B. The competitive advantage of sanctioning institutions. Science 2006; 312: 108-11.
[13]   Fletcher JA, Zwick M. The evolution of altruism: Game theory in multilevel selection and inclusive fitness. J Theor Biol 2007; 245: 26-36.
[14]   Eldakar OT, Wilson DS. Selfishness as second-order altruism. Proc Natl Acad Sci USA 2008; 105: 6982-6.
[15]   Henrich J, Boyd R, Bowles S, *et al.* In search of Homo economicus: Behavioral experiments in 15 small-scale societies. Am Econ Rev 2001; 91: 73-8.
[16]   Coe K. Art: The replicable unit – An inquiry into the possible origin of art as a social behavior. J Soc Evol Sys 1992; 15: 217-34.
[17]   Miller G. The Mating Mind. Doubleday: New York; 2000.
[18]   Griskevicius V, Cialdini RB, Kenrick DT. Peacocks, Picasso, and parental investment: The effects of romantic motives on creativity. J Pers Soc Psy 2006; 91: 63-76.
[19]   Morris D. The Human Animal. Crown Publications: Canada; 1994.
[20]   Cavalli-Sforza LL, Feldman MW. The application of molecular genetic approaches to the study of human evolution. Nature 2003; 33: 266-75.
[21]   Garrigan D, Hammer MF. Reconstructing human origins in the genomic era. Nat Rev Genet 2006; 7: 669-80.
[22]   Grinde B. The biology of visual aesthetics. J Soc Evol Sys 1996; 19: 31-40.
[23]   Jacobs GH. Primate photopigments and primate color vision. Proc Natl Acad Sci USA 1996; 93: 577-81.
[24]   Masataka N. Music, evolution and language. Devel Sci 2007; 10: 35-9.
[25]   Grinde B. A biological perspective on musical appreciation. Nord J Mus Ther 2000; 9: 18-27.
[26]   Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, Kilts CD. A neural basis for social cooperation. Neuron 2002; 35: 395-405.
[27]   Seymour B, Singer T, Dolan R. The neurobiology of punishment. Nat Rev Neurosc 2007; 8: 300-11.